TEC-0096

Vision-Based Navigation and Recognition

Azriel Rosenfeld

University of Maryland
Center for Automation Research
Computer Vision Laboratory
College Park, MD 20742-3275

August 1998

19980826 003

Approved for public release; distribution is unlimited.

Prepared for:
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714

Monitored by:
U.S. Army Corps of Engineers
Topographic Engineering Center
7701 Telegraph Road

US Army Corps of Engineers
Topographic Engineering Center

T
E
C

U.S. ARMY TOPOGRAPHIC ENGINEERING CENTER
TEC

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1998	Tech. Final April 1992-September 1995

4. TITLE AND SUBTITLE

Vision-Based Navigation and Recognition

5. FUNDING NUMBERS

DACA76-92-C-0009

6. AUTHOR(S)

Azriel Rosenfeld

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Maryland
Center for Automation Research
Computer Vision Laboratory
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Advanced Research Projects Agency
3701 N. Fairfax Drive, Arlington, VA 22203-1714

U.S. Army Topographic Engineering Center
7701 Telegraph Road, Alexandria, VA 22315-3864

19. SPONSORING / MONITORING AGENCY REPORT NUMBER

TEC-0096

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Research conducted under this contract dealt with the following topics: (1) Navigation: Navigational competencies, visual interception, robot control in navigational tasks, navigational functionalities, landmark-based localization, visibility in terrain, road following, and mobility in discrete spaces. (2) Motion analysis: flow-based methods, correspondence-based methods, and other approaches. (3) Recovery: Statistical reliability, stereo, texture, shading, and pose. (4) Invariants - both geometric and other types. (5) Human faces: Analysis of images of human faces, including feature extraction, face recognition, compression, and recognition of facial expressions. Other topics studies under this contract include fingerprints and documents, coding, diffusion, search, line fitting, matching and registration, parallel image analysis, camera motion control, and function-based object recognition.

14. SUBJECT TERMS

Navigation, motion analysis, visual recovery, invariants, face recognition, object recognition, document image understanding

15. NUMBER OF PAGES

29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

Contents

PREFACE

This report is sponsored by the Defense Advanced Research Projects Agency (DARPA) and monitored by the U.S. Army Topographic Engineering Center (TEC) under Contract DACA76-92-C-0009, titled "Vision-Based Navigation and Recognition." The DARPA Program Manager is Dr. Tom Strat, and the TEC Contracting Officer's Representative is Ms. Lauretta Williams.

The research performed under the Contract is summarized in Sections 1–6 of this report. Fifty-six technical reports were issued under the Contract. A list of these reports is given in Section 7. Numbers in brackets in Sections 1–6 refer to this list.

1 Navigation

Many issues related to navigation were investigated under this Contract. Specific areas studied included the general capabilities needed by a navigating agent; visual interception; robot control in navigational tasks; navigational functionalities; landmark-based localization; visibility in terrain; road following; and mobility in discrete spaces. Work done in each of these areas is briefly described in the following subsections.

1.1 Navigational competencies [10]

An important direction of research being pursued on the Contract has focused on the computational modeling of navigational tasks, guided by the idea of approaching vision for behavioral systems in the form of modules that are directly related to perceptual tasks. These studies led to an investigation of the problems that have to be addressed in order to obtain an overall understanding of perceptual systems, to the formulation of principles underlying the architecture of vision systems, the design and analysis of perceptual systems, and promising future research directions. The approach being pursued for understanding behavioral vision to realize the relationship of perception and action builds on two earlier approaches: the Medusa philosophy and the Synthetic approach. The resulting framework calls for synthesizing an artificial vision system by studying visual competencies of increasing complexity, and at the same time, pursuing the integration of the perceptual components with action and learning modules. It is expected that computer vision research in the future will progress in close collaboration with many other disciplines that are concerned with empirical approaches to vision, i.e. the understanding of biological vision. In particular, biological findings can motivate computational arguments that will influence future studies of computer vision.

1.2 Visual interception [17]

A visual interception system consists of one or more cameras, an agent, a target, and a mind. The mind uses information from the camera(s) in order to generate the control of the agent so that it will intercept the target. Under the traditional paradigm of considering vision as a recovery problem, visual interception is just another application of the structure from motion module: The module reconstructs the three dimensional positions and velocities of the camera, the agent, and the target, and then the information is used by a planning module to generate correct control of the agent. However, even if solutions of such three dimensional reconstruction problems are possible, they are expensive. The inherent difficulties associated with the structure from motion problem have delayed the development of real time applications, and no general visual interception system is known to exist to date.

The geometry of visual interception has been studied under the Contract. Robust solutions have been derived under the active qualitative vision paradigm. A significant finding is that the geometry of visual interception does not have to rely on depth. From the image intensity function, the *locomotive intrinsics* of the agent and the target can be obtained. Based on this relative information, a control strategy can be designed that decides in real time, and on the basis of the image intensity function, whether the velocity of the agent

should be increased or decreased at any time instant, thus guiding the agent to intercept the target. Thus, the problem of visual interception can be solved using only the spatiotemporal derivatives of the image intensity function; no correspondence is necessary. The computation is simple and can be performed in real time.

1.3 Robot control in navigational tasks [15, 16]

Traditionally, a robot's visual system is assigned the task of reconstructing the shape of the surrounding scene in the form of a depth map, which can be exploited to solve navigation problems by means of trajectory planning, control of mechanisms, etc. Unfortunately, as an overview of the state of the art in visual reconstruction reveals, it is still impossible to reliably compute depth maps, because of the fact that all shape-from-x problems are mathematically ill-posed (they have no unique solution and/or the solution is unstable). Furthermore, judging by progress made in recent years on developing methods of performing tasks such as road following and visual servoing, it appears that the depth map may not, after all, be the most suitable data representation for such tasks.

On this Contract, it has been shown that task-specific visual information with a minimum of structure is sufficient to accomplish classical navigational tasks such as obstacle avoidance and hand/eye coordination. Through these examples, it has been demonstrated that the use of vision actually simplifies the solution of robotics problems, allowing real-time control without the complex calibration procedures traditionally required.

It has been assumed in the past that the robot's trajectory is determined exclusively by the location of the goal and that the appropriate visual data can be acquired as the robot progresses toward the goal. In particular, the trajectory planning module never takes into account the needs of the visual module. It has been recognized that the stability and accuracy of visual algorithms are affected by the motion of the observer, but not much work has been done on deciding what constitutes a "good" motion. A quantitative criterion has been developed for the evaluation of the goodness of a particular action, and it has been shown how this additional layer of planning can be incorporated into the low-level control strategy, together with higher-level symbolic reasoning.

The hand/eye coordination problem can be represented, for a given pose of the observer, by the singularities of a surface, the Perceptual Control Surface (PCS). Small changes in the pose of the observer generally produce smooth deformations of the PCS. There are configurations, however, from which arbitrarily small modifications of the point of view result in profound changes in the topological nature of the PCS. These bifurcation configurations have been investigated, including the possibility of determining *a priori* displacements of the observer that can achieve a desired effect on the PCS, such as simplifying its topology or reducing the number of singularities separating the current configuration from a goal to be reached. The result of this analysis takes the form of the "family portrait" of all possible aspects of the PCS, indexed by the geometry of the manipulator and the pose of the observer relative to it. A hand/eye system is then completely coordinated—has "learned its PCS"—when a "portrait" has been matched with the experimental data gathered by the low-level controller.

1.4 Navigational functionalities [30]

A navigating agent can relate to an object in three basic ways: Avoidance (e.g., if the object is a threat or an obstacle); interception (e.g., if the object is prey or food); and reference (e.g., if the object can be used as a landmark). These classes of object functionalities have been studied in the case in which the agent is a simple corridor-cleaning robot that uses the walls (and wall-floor junctions) as references in following the corridor; treats independently moving objects as threats; treats large stationary objects as obstacles; and treats small stationary objects as "prey" (trash to be swept up).

1.5 Landmark-based localization [20, 26]

A method for localization and positioning in an indoor environment has been developed. *Localization* is the act of recognizing the environment, and *positioning* is the act of computing the exact coordinates of a robot in the environment. The method is based on representing the scene as a set of 2D views and predicting the appearances of novel views by linear combinations of the model views. The method accurately approximates the appearance of scenes under weak perspective projection. Analysis of this projection, together with experimental results, demonstrates that in many cases this approximation is sufficient to accurately describe the scene. When weak perspective approximation is invalid, either a larger number of models can be acquired or an iterative solution to account for the perspective distortions can be employed.

The method has several advantages over other approaches. It uses relatively rich representations; the representations are 2D rather than 3D; and localization can be done from only a single 2D view. The same principal method is applied for both the localization and positioning problems, and a simple algorithm for *repositioning*, the task of returning to a previously visited position defined by a single view, is derived from this method.

Navigation also can be studied in a graph-structured framework in which the navigating agent (which can be regarded as a point robot) moves from node to node of a "graph space." The robot can locate itself by the presence of distinctively labeled "landmark" nodes in the space. For a robot navigating in Euclidean space, visual detection of a distinctive landmark provides information about the direction to the landmark, and allows the robot to determine its position by triangulation. On a graph, however, there is neither the concept of direction nor that of visibility. Instead, one can assume that a robot navigating on a graph can sense the distances to a set of landmarks.

Evidently, if the robot knows its distances to a sufficiently large set of landmarks, its position on the graph is uniquely determined. This suggests the following problem: given a graph, what is the smallest number of landmarks needed and where should they be located, so that the distances to the landmarks uniquely determine the robot's position on the graph? This smallest number is known as the metric dimension of the graph. It can be computed in linear time for trees and grid graphs. Computing it is computationally intractable for arbitrary graphs, but it can be approximated in polynomial time within a factor of $O(\log n)$.

1.6 Visibility in terrain [37, 38]

Two classes of parallel algorithms have been investigated under the Contract for point-to-region visibility analysis on digital terrain models: ray-structure-based methods and propagation-based methods. A new propagation-based algorithm has been developed to avoid problems that commonly arise in simple propagation-based algorithms. The performance and characteristics of the two classes of algorithms have been compared; the sources of uncertainty in visibility computation, and the importance of taking uncertainty into consideration, also have been investigated. Different methods of representing the uncertainty have been studied, including Monte Carlo simulation, analytic estimation, and some simple heuristic indicators. Experiments show that these simple heuristic indicators can be used for efficient coarse classification of the likelihood of point intervisibility.

Several parallel algorithms for visibility and triangulation problems have been designed and evaluated using a mixture of theoretical and practical criteria. Most of these algorithms have been implemented and experimental results obtained for them. In particular, a parallel algorithm has been developed for computing region-to-region visibility on polyhedral terrain models. Its complexity is $O(\log^2 n)$ time and $O((n\alpha(n)+k)\log n)$ operations on a Concurrent-Read, Exclusive-Write Parallel Random-Access Machine (CREW PRAM), where n and k are the input and output sizes, respectively, and α is the inverse of Ackermann's function. The vertex-ray method for computing approximate visibility on polyhedral terrain also has been studied; it is suitable for data-parallel implementation.

A method of triangulating a terrain surface in parallel also has been developed. The method is able to construct either a hierarchical triangulation or a Delaunay triangulation from a digital terrain model. It refines the triangulation iteratively to the desired precision. A parallel algorithm for 3D Delaunay triangulation also has been designed, and techniques have been developed to cope with several important issues such as load-balancing and robustness.

1.7 Road following [31]

Recently, connectionist architecture approaches have been employed to solve the autonomous visual road-following problem. Carnegie-Mellon University implemented a feed-forward multilayer perceptron (FMLP) network controller called ALVINN (Autonomous Land Vehicle in a Neural Network) that very successfully performed visual road-following. ALVINN, however, experienced problems with primitive road feature retention, spurious control signals in the presence of structured noise, and learning anomalous driving situations. An alternative connectionist architecture, called a Radial Basis Function (RBF) network, was developed under the Contract for visual road-following. A controller based on each network architecture was built, and the performance of the two controllers was evaluated using a driving simulator. The FMLP controller experienced the same problems on the driving simulator that ALVINN experienced in real road-following. The RBF controller did not experience any of the problems that the FMLP network experienced, and did not suffer any negative side-effects.

1.8 Mobility in discrete spaces [21]

Most previous theoretical work on motion planning has addressed the problem of path planning for geometrically simple robots in polyhedral terrains. Under the Contract, motion planning in a graph-structured space has been studied. Conditions have been established under which a robot having a particular graph structure can move from any start configuration to any goal destination in a graph-structured space.

2 Motion analysis

Several approaches were taken under the Contract to solving the problem of deriving motion information from image sequences, including methods based on (normal) image flow and methods based on feature point correspondences, as well as various special methods and analyses. This work is summarized in the following subsections.

2.1 Flow-based methods [8, 9, 11, 12]

Because of the aperture problem, the only general unambiguous motion measurement in images is normal flow—the projection of image motion on the gradient direction. A monocular observer can estimate its 3D motion relative to the scene by using normal flow measurements in a global and mostly qualitative way. This problem can be addressed through a search technique. By checking constraints imposed by 3D motion parameters on the normal flow field the possible space of solutions is gradually reduced. In the four modules that comprise the solution, constraints of increasing restriction are considered, culminating in testing every normal flow value for its consistency with a set of motion parameters. The fact that motion is rigid defines geometric relations between certain values of the normal flow field. The selected values form patterns in the image plane that are dependent on only some of the motion parameters. These patterns, which are determined by the signs of the normal flow values, are searched for in order to find the axes of translation and rotation. The third rotational component is computed from normal flow vectors that are only due to rotational motion. Finally, by looking at the complete data set, all solutions that cannot give rise to the given normal flow field are discarded from the solution space.

The human eye is different from existing electronic cameras because it is not equipped with a uniform resolution over the whole visual field. It has the fovea near the optical axis where the resolution (over a one degree range) is higher by an order of magnitude than that in the periphery. With a small fovea in a large visual field, it is not surprising that the human visual system has developed mechanisms, usually called saccades or pursuits, for moving the fovea rapidly. It is important to understand both the structure and function of eye movements in the process of solving visual tasks. In other words, how does this particular ability of humans and primates to fixate on environmental points in the presence of relative motion help their visual systems in performing various tasks? In a more formal setting, the following problem can be posed: Consider an anthropomorphic active vision system, that is, a pair of cameras resting on a platform and controlled through motors by a computer that has access to the images sensed by the cameras in real time. The platform can move

freely in the environment. If this machine can fixate on targets that are in motion relative to it, can it perform visual tasks in an efficient and robust manner? It turns out that such an active observer can solve the problems of 3D motion estimation, egomotion recovery and estimation of time to contact in a very efficient manner. The algorithms for solving these problems are robust and of a qualitative nature and employ as input only the spatiotemporal derivatives of the image intensity function (i.e. they make no use of correspondence or optic flow). Fixation is achieved through camera rotation. This amounts to a change of the input (motion field) in a controlled way. From this change additional information is derived making the previously mentioned navigational problems easier to solve. A system possessing gaze control capabilities also can successfully address other problems, such as figure-ground segmentation, stereo fusion, visual servoing for manipulatory tasks, and determination of relative depth. These findings demonstrate that gaze control is an important principle of active vision.

In general, image displacement fields—optical flow fields, stereo disparity fields, normal flow fields—produced by rigid motion possess a global geometric structure that is independent of the scene in view. Motion vectors of certain lengths and directions are constrained to lie on the imaging surface at particular loci whose location and form depend solely on the 3D motion parameters. If optical flow fields or stereo disparity fields are considered, then equal vectors are shown to lie on conic sections. Similarly, for normal motion fields, equal vectors lie within regions whose boundaries also constitute conics. By studying various properties of these curves and regions and their relationships, a characterization of the structure of rigid motion fields can be given. A concept underlying the global structure of image displacement fields has been researched and developed under the Contract. This concept gives rise to various constraints that could form the basis of algorithms for the recovery of visual information from multiple views.

Recent experiments in primates have revealed that, in certain areas of the brain, there exist cells with large receptive fields responding to patterns of visual motion. Geometric considerations described here suggest that these patterns could be spatial arrangements or gratings, i.e., aggregations of orientations along which retinal motion information is estimated. The exact form of the gratings is defined by the shape of the retina; for a planar retina they are radial lines, concentric circles, as well as elliptic and hyperbolic curves, while for a spherical retina they become longitudinal and latitudinal circles for various axes. Considering retinal motion information computed normal to these gratings, patterns are found that have encoded in them subsets of the 3D motion parameters. The importance of these patterns is, first, that they depend on only the 3D motion and not on the scene in view, thus providing globally a separation of the effects of 3D motion and structure on the image motion; and, second, that they are founded upon easily derivable image measurements—they do not use exact retinal motion measurements, such as optical flow, but only the sign of the image motion along a set of directions defined by the gratings. Based on the geometrical findings, a computational theory has been developed that shows that the problem of estimating the egomotion of a system can be efficiently solved through pattern matching. This theory or variations of it may be implemented in nature; this can be established through experiments in the neurosciences.

2.2 Correspondence-based methods [49, 50, 51, 56]

A general motion correspondence algorithm has been developed. The basic problem of motion correspondence is simply to track the same features over consecutive frames, a challenging problem when camera motion is significant. In general, feature displacement over consecutive frames can be approximately decomposed into two parts: The displacement caused by camera motion, which can be compensated by image rotation, scaling, and translation; and the displacement caused by object motion and/or perspective projection. The displacement caused by camera motion is usually much larger and more irregular than the displacement caused by object motion and perspective distortion. A two-step approach has been developed to solve this problem. First, the motion of the camera is estimated using a recently developed image registration algorithm. Then consecutive frames are transformed to the same coordinate system; this reduces the feature correspondence problem to that of tracking moving objects using a still camera. A method for subpixel accuracy matching also has been introduced to reduce feature drift over a long sequence. The approach results in a robust and efficient algorithm. Good experimental results have been obtained on several image sequences.

A simple but robust model based approach has been developed for estimating the kinematics of a moving camera and the structure of the objects in a stationary environment using long, noisy, monocular image sequences. Both batch and recursive algorithms have been formulated and the problems that arise because of occlusion have been addressed. The approach is based on representing the constant translational velocity and constant angular velocity of the camera motion using nine rectilinear motion parameters, which are 3D vectors of the position of the rotation center, linear and angular velocities. The structure parameters are 3D coordinates of the salient feature points in the inertial coordinate system. Thus, the total number of parameters to be estimated is $3M + 7$, where M is the number of feature points. The image plane coordinates of these feature points in each frame are first detected and then matched over the frames. These noisy image coordinates serve as the input to the algorithms. Because of the nonlinear nature of perspective projection, a nonlinear least squares method has been formulated for the batch algorithm, and a conjugate gradient method is then applied to find the solution. A recursive method using an Iterated Extended Kalman Filter (IEKF) for incremental estimation of motion and structure also has been developed. Since the plant model is linear in this formulation, closed form solutions for the state and covariance transition equations can be directly derived. Experimental results have been obtained for simulated imagery as well as several real image sequences. For the simulated imagery, the bias and the variance of the estimators have been analyzed by the Monte Carlo method; while for the real imagery, the ground truth of the motion and structure parameters (when available) have been compared with the computed results.

An algorithm also has been developed for estimating the kinematics of a moving camera(s) and the structure of the objects in a stationary environment using long, noisy stereo image sequences. The kinematics of the camera(s) is modeled to be a constant translational and rotational motion using nine rectilinear parameters. The total number of parameters to be estimated is $3M+8$, where M is the number of feature points. Problems because of occlusion and non-uniform temporal sampling have been addressed and conditions for uniqueness of the motion and structure parameters have been formulated. Image plane coordinates of

automatically extracted feature points over the stereo sequence are used as inputs. A real-imagery based experiment also has been performed.

In the analysis of visual motion from long, noisy image sequences, both the moving camera-stationary scene and fixed camera-moving object cases have been studied. The task involved is the estimation of motion and structure parameters from 2D image coordinates. Under the central projection model, a kinematic model based approach has been developed to relate the time evolution of the parameters to the 2D image plane feature coordinates. Two approaches, batch and recursive, have been developed to solve this highly nonlinear problem. In the batch approach, a cost function is defined, and a conjugate gradient method is applied to find all the parameters that minimize this cost function. In the recursive approach, an IEKF is used to propagate and update the state variables.

Based on the assumption that the frame sampling rate is high enough, and thus, the motion is smooth over a short period of time, only the first order motion parameters are used to model the camera (or object) motion. Therefore, in situations where departures from the assumed motion model exist, the batch approach provides rough estimates over the first few frames; while deviations from the assumed model are handled using the tracking ability of the Kalman filter. For all the cases considered, the standard rotation matrix is used to represent the rotational motion instead of the commonly used quaternions. This results in a simple linear plant model in the recursive method. Closed form solutions for the state and covariance transition equations are obtained without the use of the time-consuming numerical integration step.

In the monocular case, since motion and structure parameters are estimated at the same time, the processing of the batch algorithm over many frames would be time-consuming. The batch algorithm instead can be used to give rough estimates for all the parameters over the first few frames. Then, these values can be used as initial guesses for the recursive algorithm. In order to prevent the Kalman filter from diverging, a good initial guess for the covariance matrix is often necessary. This can be done by computing the Cramér-Rao lower bounds for all the parameters and using these bounds as the initial values of the covariance matrix.

In the binocular case, since the 3D structures can be roughly estimated from the first pair of left and right images using the classic stereo triangulation method, both the batch and recursive algorithms converge more easily than in the monocular case. A proof of the uniqueness of the parameters also has been given for the binocular case. It has been shown that three noncollinear feature points over three consecutive frames contain all the information needed for motion and structure estimation. Based on this proof, the rotational parameters can be estimated first using a deterministic method. Subsequently, all the other parameters can be estimated using a linear algorithm, leading to a much faster implementation.

In real experiments, the accuracy of the algorithms depends heavily on the calibration of the camera being used. The problem of camera calibration has been addressed in a straightforward manner. If the 3D structures of some feature points are available, then based on a simple batch algorithm, the image center and the field of view can be roughly estimated to improve the performance of the estimation process.

Several real image sequences have been carefully tested. The ground truth, when available, has been compared to the estimates. For most of the real sequences, the inputs to the algorithms, which are 2D image coordinates, are all automatically detected and matched

over frames. Despite the presence of various types of noise as well as the occlusion problem, the long frame sequence based methods yielded good results, although as few as four feature points were used.

2.3 Other approaches [3, 4, 19, 39]

A noise robust algorithm for computing 3D motion from images has been developed for the feature-based two-view problem—computing the depths of the feature points and the camera motion from correspondences of feature points between two images. The decomposability condition, uniqueness of the solution, direct optimization, and the "critical surface" also have been derived; planar surface motion has been treated separately. A noise robust algorithm also has been developed for optical flow. The decomposability condition, uniqueness of the solution, direct optimization, the "critical surface," and relationships to the algorithm for finite motion all have been investigated; planar surface optical flow has been treated separately.

Situations involving navigation among maneuvering agents are critical for the study of visual guidance of autonomous vehicles. The general case of motions with polynomial laws in their translational component has been addressed, and a class of temporal parameters (TP) has been defined that is relevant for navigation, and enables qualitative descriptions of the observed depth trajectories. The parameters have been shown to be visually recoverable. Examples of these parameters include the Time to Collision (TTC) and the Time to Synchronization (TTS) (the time until the observer achieves the same velocity as a moving object, relevant for docking or platooning maneuvers). Recoverability of these parameters leads to an equivalent temporal representation of visual information. Results on recoverability are specialized to lower order motions and the recovery of TTC and TTS for arbitrarily smooth laws. Computations using direct and feature-based methods have been carried out. A scheme for addressing model order determination, collision detection and temporal parameter estimation also has been formulated and tested, and experimental results on synthetic and real images have been obtained.

Transform methods have been applied to the analysis of dynamic image sequences and to the characterization of image motion. Image motion resulting from arbitrary 3D camera translation is conveniently analyzed in the Mellin Transform (MT) domain associated with space as well as time dimensions, resulting in the desired separation of the generalized spectrum into a structural component corresponding to the spatial MT of the static image and an MT component depending on the image motion itself (a motion support). This result has potential implications for the recovery of image motion from integral image brightness measurements. In particular, the effects of "nearness" to an imaged object and TTC on the resulting MT spectral motion support has been investigated. Conversely, the recovery of TTC from MT spectral analysis along the time and space directions has been studied, and different cases of Mellin parameters have been examined.

Within the computer vision community the paradigm of active vision has attracted increasing attention over the last few years. Most of the related work can be roughly subdivided into more conceptual or theoretical work (linearization of problems by active sensor movements, novel algorithms, proving existence and uniqueness of solutions, etc.) and more

practical work (the construction of active head systems and the implementation of simpler algorithms in real time). Almost all of the theoretical work related to motion is based on the instantaneous motion field model. An active camera, however, acquires images at discrete time steps, resulting in finite observer motion between consecutive frames. The crude approximation of infinitesimal movements between the frames is not adequate if information is integrated over time, as, for example, under fixation or for robust determination of structure from motion. Fixation is a basic behavioral capability of a moving observer, facilitating, for instance, navigation using one or more landmarks (dead reckoning) and shape from motion determination during ego-motion.

Mathematical tools have been developed to deal with vision-related problems under finite kinematics for applications that are by nature discrete in time. A general fixation constraint has been derived using finite kinematics. The pros and cons of several distinct kinematic sensor models (e.g., combinations of the rotational degrees of freedom of the active visual sensor) have been investigated, and closed-form solutions for the inverse kinematic problem have been given in all cases. An additional constraint can be imposed on the observer motion to yield a unique solution for a redundant manipulator. General time-dependent formulas have been derived for the projected trajectories, the motion field, the projected displacement vector field, and the trajectories of the epipole or FOE in the image plane. Robust algorithms have been formulated for the recovery of the observer trajectory, the direction of motion, and a novel way to compute the optical flow field from image data obtained by a fixating observer.

3 Recovery

Other research on 3D vision conducted under the Contract has dealt with the statistical reliability of 3D interpretation techniques; with stereo; with analysis of two- and three-dimensional texture; with shape from shading; and with pose estimation. This work is briefly described in the following subsections.

3.1 Statistical reliability [18]

The reliability of 3D interpretations computed from images can be analyzed in statistical terms by employing a realistic model of image noise. First, the reliability of edge fitting is evaluated in terms of image noise characteristics. Then, the reliability of vanishing point estimation is deduced from the reliability of edge fitting. The result is applied to focal length calibration, and an optimal scheme derived in such a way that the reliability of the computed estimate is maximized. The confidence interval of the optimal estimate also is computed. Also considered is the reliability of fitting an orthogonal frame to three orientations obtained by sensing. Finally, statistical criteria have been derived for testing edge groupings, vanishing points, focuses of expansion, and vanishing lines.

3.2 Stereo [6]

Since Baker introduced the use of dynamic programming for stereo matching, extensive research has focused on this idea, including work by Arnold, Ohta and Kanade, and Lloyd.

While previous researchers have pointed out that such methods are suitable for parallel processing, there have been no attempts to implement the dynamic programming stereo matching algorithm on a parallel machine. A massively parallel implementation of Baker's dynamic programming algorithm has been formulated; this implementation can use many processors per scanline, compared to a naive approach of one processor per scanline. This is important because typical images contain 256 to 1024 scanlines, while massively parallel machines can have many more processors. A method of handling inter-scanline inconsistencies that is very well suited for parallel implementation also has been introduced. The method increases the total amount of processing needed to solve the stereo matching problem by only a small fraction. Parallel implementations of both the dynamic programming algorithm and the inter-scanline inconsistency correction algorithm have been fully analyzed. Timing results show that on a 16K processor Connection Machine the entire algorithm requires from only 1 second for simple 512×512 images, to 12 seconds for complicated ones.

3.3 Texture [14, 41, 42, 43]

A new method of texture analysis and classification has been developed based on a local center-symmetric covariance analysis, using Kullback (log-likelihood) discrimination of sample and prototype distributions. This analysis features generalized, invariant, local measures of texture having center-symmetric patterns, which is characteristic of most natural and artificial textures. Two local center-symmetric auto-correlations, with linear and rank-order versions (SAC and SRAC), have been introduced, together with a related covariance measure (SCOV) and variance ratio (SVR). All of these are rotation-invariant, and three are locally grey-scale invariant, robust measures. In classification experiments, their discriminant information has been compared to that of Laws' well-known convolutions, which have specific center-symmetric masks. The new covariance measures achieved very low classification error rates despite their abstract measure of texture pattern and grey-scale.

Basic classes of three-dimensional textures also have been studied under the Contract. In particular, the distribution of leaves in a tree crown has been modeled, and two statistical properties of such distributions have been studied: the probability of seeing through the leaves, and the distribution of leaf gray levels.

Formally, the model deals with three-dimensional textures composed of opaque planar texels uniformly distributed over a volume of space. For simple assumptions about the shapes of the texels and their distribution of orientations, the probability of seeing through a given-thickness volume of the texture can be estimated; these estimates have been confirmed using synthetic examples. The probability is quite insensitive to texel shape (for a given average texel area), but is more sensitive to the distribution of texel slants, since the slant of a texel affects its subtended area. For example, expected texel slants tend to be high for textures composed of "leaves" whose stems conform to a standard tree branching model. On the other hand, in real scenes containing falling disks ("snowflakes"), it was found that the disks had about the same average slant as texels whose distribution of orientations is uniform.

The gray level histograms of images of such textures also have been studied under illumination by a compact light source. Simple models can be used to describe the variation of such histograms with light source direction. In fact, the variation of real plant histograms

with light source direction resembles that of synthetic histograms generated using a Phong-type reflectance model and a uniform texel orientation model, and ignoring transmittance, interreflection, and shadows.

3.4 Shading [35] and pose [25]

An improved shape from shading (SFS) algorithm has been developed that is an extension of the recently published algorithm by Zheng and Chellappa. A markedly more accurate estimate of the azimuth of the illumination source is obtained. Depth reconstruction also is improved by using a new set of boundary conditions and adapting a more sophisticated technique for hierarchical implementation of the SFS algorithm. Errors at the boundaries of images and in rotation of the reconstructed images have been corrected. Results on synthetic and real images have been obtained.

A new method has been developed for the computation of the position and orientation of a camera with respect to a known object, using four or more *coplanar* feature points. Starting with the scaled orthographic projection approximation, this method iteratively refines up to two different pose estimates, and provides an associated quality measure for each pose. When the object's distance to the camera is large compared with the object's extent along the direction of the optical axis, or when the accuracy of feature point extraction is low because of image noise, the two quality measures are similar, and the two pose estimates are plausible interpretations of the available information. In contrast, known methods using a closed form pose solution for four coplanar points are not robust for distant objects in the presence of image noise because they provide only one of the two possible poses and may choose the wrong pose.

4 Invariants

Invariant descriptors are useful for object recognition because they are independent of the viewpoint from which the image was taken. It also is possible to define invariants for small object deformations, or for physical qualities related to the image formation process. Research on invariants conducted under the Contract is described in the following subsections.

4.1 Geometric invariants [27, 45, 48]

A new and more robust method of obtaining local projective and affine invariants has been developed. These shape descriptors are useful for object recognition because they eliminate the search for the unknown viewpoint. Being local, these invariants are much less sensitive to occlusion than the global ones used by other researchers. The basic ideas are employing an implicit curve representation without a curve parameter, thus increasing robustness; and using a canonical coordinate system which is defined by the intrinsic properties of the shape, independent of any given coordinate system, and is thus invariant. Several configurations have been treated: a general curve without any correspondence, and curves with known correspondences of one or two feature points or lines. The method has been applied to real images without the use of a curve parameter by fitting an implicit polynomial to a general

curve in a neighborhood of each curve point. Experimental results for various 2D objects in 3D space have been obtained.

Image databases are conceptually much harder to deal with than conventional databases because the information that they contain consists of images, rather than alphanumeric entities. Images can be fuzzy and distorted, and they depend on the point of view from which the object is seen. Characteristics of the images which are invariant to changes in the viewpoint are presented. These characteristics can be stored as "signatures" for the objects in an atlas database, thereby permitting efficient retrieval and matching (partial or total) regardless of the viewpoint. Invariant signatures are useful because image matching is very slow in high population image databases. They also can be indexed easily using current database technology (e.g., the B-tree). Strategies for the processing of queries in such an environment have been formulated.

4.2 Other types of invariants [28, 46, 47]

The general invariance concept plays an important role in object recognition. There is extensive literature, both classical and recent, on projective invariance. Invariants help solve major problems of object recognition. For instance, different images of the same object often differ from each other because of the different viewpoints from which they were taken. To match the two images, common methods need to find the correct viewpoint; this is a difficult problem that can involve search in a large parameter space of all possible points of view and/or finding point correspondences. Geometric invariants are shape descriptors, computed from the geometry of the shape, that remain unchanged under geometric transformations such as changing the viewpoint; thus, they can be matched without a search. Deformations of objects are another important class of changes for which invariance is useful.

Object recognition means not only recognizing a particular shape, but recognizing a class of shapes that are related to each other in some way. For example, two shapes can be regarded as related if one of them can be deformed into the other. The deformation must belong to some predefined set of deformations; it should not be too general. Invariants have been developed for quasi-affine deformations, i.e. transformations that are approximately linear but also have small non-linear components. Shape descriptors have been defined that are "quasi-invariant" to these deformations, and they have been used to recognize classes of real objects.

Invariants also can be formulated that concern the physical processes that form images, involving shading, IR, radar, sonar, etc. The image formed by such a process depends on many variables in addition to the geometry, such as the characteristics of the lighting or other incident radiation, the imaging system, etc. Most of these variables are not known in advance, so the recovery of shape is difficult. The problem could be greatly simplified if it were possible to find invariants of the situation, namely quantities that stay unchanged when some of the unknown variables change. Known methods of mathematical physics can be applied to finding invariants of physical imaging processes. These methods take advantage of various symmetries, which can be part of a model-based approach to recognition. The approach has been illustrated for the case of the shape from shading problem, but the methods have much wider applicability.

5 Human faces

Research was initiated on the Contract dealing with problems relating to the recognition of human faces, the compression of images involving facial motions, and the recognition of facial expressions. This work is described in the following subsections.

5.1 Faces: Features, recognition, compression [5, 36, 40, 53]

An approach to labeling the components of human faces from range images has been developed. The components of interest are those that humans usually find significant for recognition. To cope with the non-rigidity of faces, an entirely qualitative approach has been used. The preprocessing stage employs a multi-stage diffusion process to identify convexity and concavity points. These points are grouped into components and qualitative reasoning about possible interpretations of the components is performed. Consistency of hypothesized interpretations is carried out using context-based reasoning. Experimental results have been obtained on real range images of faces.

Methods also have been developed for the segmentation and identification of human faces from grey scale images with clutter. The segmentation process uses the elliptical structure of the human head. It uses the information present in the edge map of the image, and through some preprocessing, separates the head from the background clutter. An ellipse is then fitted to mark the boundary between the head region and the background. The identification procedure finds feature points in the segmented face through a Gabor wavelet decomposition and performs graph matching. The segmentation and identification algorithms have been tested on a database of 48 images of 16 persons with good results.

Very-low bandwidth video-conferencing, which is the simultaneous transmission of speech and pictures (face-to-face communication) of the communicating parties, is a challenging application requiring an integrated effort of computer vision and computer graphics. A simple approach to video-conferencing has been developed relying on an example-based hierarchical image compression scheme. In particular, consideration has been given to the use of example images as a model, the number of required examples, faces as a class of semi-rigid objects, a hierarchical model based on decomposition into different time-scales, and the decomposition of face images into patches of interest. Approaches to face recognition that are relevant to this work also have been evaluated. Algorithms have been designed and evaluated for image processing and animation, including an automatic algorithm for pose estimation and normalization. Experiments suggest interesting estimates of necessary spatial resolution and frequency bands. Algorithms for finding the nearest neighbors in a database for a new input have been reviewed and compared, and a generalized algorithm has been developed for blending patches of interest in order to synthesize new images. Extensions to image sequences have been formulated together with possible extensions based on the techniques of Beymer, Shashua and Poggio for interpolating between example images. The integration of these algorithms can be used to define a simple model-based video-conferencing system.

A critical survey of the literature on human and machine recognition of faces has been conducted. Machine recognition of faces has several applications ranging from static matching of controlled photographs, as in mugshot matching and credit card verification, to surveillance

using video images. These applications have different constraints in terms of the complexity of their processing requirements and thus present a wide range of technical challenges. Over the last twenty years researchers in psychophysics, neural sciences and engineering, image processing, analysis, and computer vision, have investigated a number of issues related to face recognition by humans and machines. The ongoing research activities have been given renewed emphasis over the last five years. The existing techniques and systems have been tested on different sets of images of varying complexities, but very little synergism exists between studies in psychophysics and the engineering literature. Most importantly, no evaluation or benchmarking studies exist using large databases with the image quality that arises in law enforcement/commercial applications.

Different applications of face recognition in the law enforcement and commercial sectors have been reviewed. Special constraints that are present in these applications have been pointed out, and a brief overview of the literature on face recognition in the psychophysics community has been prepared. A detailed overview has been conducted of more than twenty years of research done in the engineering community. Techniques for segmentation/location of the face, feature extraction and recognition have been reviewed. Global transform and feature based methods using statistical, structural and neural classifiers have been summarized, and a brief summary of recognition methods, using face profiles and range image data, also has been given.

Real-time recognition from video images acquired in a cluttered scene, such as an airport, is probably the most challenging face recognition problem. Not much work has been reported on this problem; however, several existing technologies in the image understanding literature could potentially impact it.

Given the numerous theories and techniques that are applicable to face recognition, it is clear that evaluation and benchmarking of these algorithms is crucial. Relevant issues include data collection, performance metrics and evaluation of systems and techniques.

5.2 Facial expressions [2, 32, 54]

An approach has been developed for the analysis and representation of facial dynamics for recognition of facial expressions from image sequences. The algorithms use optical flow computation to identify the directions of rigid and non-rigid motions that are caused by human facial expressions. A mid-level symbolic representation motivated by linguistic and psychological considerations has been developed. Recognition of six facial expressions, as well as eye blinking, has been demonstrated on a large set of image sequences.

A radial basis function network architecture that learns the correlation between facial feature motion patterns and human emotions also has been developed. A hierarchical approach is used, that at the highest level identifies emotions, at the mid level determines motions of facial features, and at the lowest level recovers motion directions. Individual emotion networks were trained to recognize the "smile" and "surprise" emotions. Each network was trained by viewing a set of sequences of one emotion for many subjects. The trained neural network was then tested for retention, extrapolation and rejection ability. Success rates were about 88% for retention, 73% for extrapolation, and 79% for rejection.

Local parametric models of image motion have been used for recovering the non-rigid and

articulate motion of human faces. Parametric flow models (for example affine) are popular for estimating motion in rigid scenes. Within local regions in space and time, such models not only accurately model non-rigid facial motions, but also provide a concise description of the motion in terms of a small number of parameters. These parameters are intuitively related to the motion of facial features during facial expressions, such as anger, happiness, and surprise. These facial expressions can be recognized from the local parametric motions in the presence of significant head motion. A large set of experiments on this problem has been performed.

6 Techniques and applications

Other computer vision applications investigated on the Contract include the recognition of fingerprints and the recovery of information from handwritten documents. A variety of basic tools and techniques also have been studied, including region-adaptive image coding; diffusion processes; search; line fitting; matching and recognition; parallel image analysis; camera motion control; and function-based object recognition. This work is briefly reviewed in the following subsections.

6.1 Fingerprints and documents [7, 44]

New algorithms for the enhancement and minutiae extraction of fingerprint images have been developed. For enhancement, two kernels have been designed. First, an eigenfilter is derived as the solution to a constrained optimization problem, and second, the bandwidth of this filter is expanded in the Fourier domain to ensure robustness against artifacts. A new algorithm for the localization of minutiae in the fingerprint image also has been developed. A Karhunen-Loeve decomposition is used as a measure of structural integrity to localize features. A Fourier approximation to the KL decomposition has been used for practical reasons. Finally, the feature detection scheme of Manjunath and Chellappa has been implemented, and its effectiveness in detecting regions of high curvature and line endings, which are an identifying characteristic of minutiae, has been studied.

Many document image understanding problems require a more comprehensive examination of document features than is typically deemed necessary for recognition tasks. These problems require a detailed analysis of stroke and sub-stroke features in the document image with the goal of obtaining information about the environment or process which created the document, and establishing a context for document understanding. In particular, the concept of *recovery* can be extended into the document domain. A "stroke platform" representation has been developed which establishes a verifiable "link to the pixels," and its usefulness for recovery tasks has been demonstrated. This representation makes it possible to overcome many of the problems associated with the rapid, irreversible abstraction associated with traditional document processing methods and provides the basic framework for an analysis of handwritten documents. By obtaining a detailed description of the document and its properties, it is possible to establish a context for analysis and validate assumptions about the domain. Several document image understanding problems have been treated. The stroke platform has been successfully used for the problem of interpreting and reconstructing junctions and endpoints. A model for instrument grasp has been developed and

the recovery of pressure patterns from a document has been demonstrated. Methods have been developed for recovering temporal information from static images of handwriting. Various problems associated with processing form documents also have been studied. Finally, the detailed analysis philosophy has been extended to demonstrate its feasibility in many document domains.

6.2 Coding [22]

Two region adaptive image coding schemes guided by different criteria have been developed. The first scheme, Region Adaptive Subband Image Coding (RA-SBIC), is based on rate-distortion theory. The major concern is minimizing the mean squared error between the original image and the coded image for a given number of bits. The second scheme, Segmentation Based Image Coding (SB-IC), incorporates ideas from Human Visual System (HVS) models. An image resembling the original image is generated using a multi-component decomposition of the original image.

In RA-SBIC, the input image is decomposed into a number of image subbands and their statistical properties are analyzed. It is observed that the amount of energy in image subbands increases towards lower frequency subbands and, more importantly, towards image edges. It also is found that the directionality of the energy distribution is highly dependent on the orientation of the edges. Based on the energy distribution, arbitrarily shaped regions are extracted in each subband and entropy-constrained quantizers using the generalized Gaussian distribution for modeling the image subbands are employed. The problem of determining an optimal subband decomposition among all the possible decompositions also is addressed. Experimental results show that visual degradations are negligible at a bit rate of 1.0 bits/pel and that reasonable quality images are obtainable up to rates as low as 0.25 bits/pel.

RA-SBIC also has been applied to video coding. For motion compensation of two consecutive frames, a feature matching algorithm is employed. Motion compensated frame differences are divided into three regions called stationary background, moving objects, and newly emerging area. Different quantizers are used for the different regions.

SB-IC mimics the processing of the contours and textures in the HVS. Using both uniform and textured region extraction algorithms, the input image is segmented. Textured regions are reconstructed using 2D noncausal Gaussian-Markov random field models. Uniform regions are reconstructed using polynomial expansions. Images of reasonable quality are obtained up to rates of 0.1 bits/pel.

6.3 Diffusion [52, 55]

A multi-stage physical diffusion process can be used for various purposes in range image processing. The input range data is interpreted as occupying a volume in 3D space. Each diffusion stage simulates the process of diffusing the boundary of the volume into the volume. The outcome is a procedure that provides for a combined discontinuity detection and segmentation into shape coherent regions. The process has been analyzed on noise-free and noisy step, roof, and valley edges. It also has been applied to real range images of indoor scenes, outdoor-terrain data, isolated real objects, and human faces.

The diffusion process also provides a new approach to image interpolation and metamorphosis. This approach is based on a scale space created by diffusing the difference function of the source and the goal images. This formulation of the problem makes it possible to minimize the need for human intervention in the selection of features in an application such as image metamorphosis. The smooth transitions are accompanied by a moderated blurring that is useful in displaying the process of metamorphosis. The proposed approach has been demonstrated on color images and range images of human faces, and on a motion image sequence as a method of enhancing animation.

6.4 Search [1]

Finding nearest neighbors is one of the most fundamental problems in computational geometry, with applications to many areas such as pattern recognition, data compression and statistics. The nearest neighbor problem is: given a set of n points in d-dimensional space, $S \subset E^d$, and given a query point $q \in E^d$, find the point of S that minimizes the Euclidean distance to q. It is assumed that d is a constant, independent of n.

Efficient algorithms are known for computing nearest neighbors in low dimensional spaces. But, as dimensionality increases, the difficulty of solving the nearest neighbor problem, either in time or space, seems to grow rapidly. It can be shown, however, that if one is willing to consider approximate nearest neighbors, rather than exact nearest neighbors, it is possible to achieve efficient asymptotic performance for both space and query time, irrespective of input distribution.

Several practical algorithms have been developed for performing nearest neighbor searching in high dimensions. These algorithms draw on data structures such as the k-d tree and the neighborhood graph. Empirical analyses of these algorithms have been shown in the context of vector quantization, which is a technique used in the compression of speech and images. The results show that these algorithms achieve massive reductions in running time while suffering little loss in performance.

Finally, the complexity of the k-d tree algorithm has been analyzed for the uniform distribution, taking into account the effect of the boundary. This provides a more accurate analysis for realistic instances in high dimensions, where boundary effects are significant.

6.5 Line fitting [24]

The conventional ordinary least squares (OLS) method of fitting a line to a set of data points is notoriously unreliable when the data contains points coming from two different populations: randomly distributed points ("random noise"), and points correlated with the line itself (e.g., obtained by perturbing the line with zero-mean Gaussian noise). Points which lie far away from the line (i.e., "outliers") usually belong to the random noise population; since they contribute the most to the squared distances, they skew the line estimate from its correct position. An analytic method of separating the components of the mixture has been developed. Unlike previous methods, this approach leads to a closed form solution. Applying a variant of the method of moments (MoM) to the assumed mixture model yields an analytic estimate of the desired line. Good experimental results have been obtained by this method.

6.6 Matching and registration [13, 33]

Correlation-based matching methods are known to be very expensive when used on large image databases. Ways of speeding up correlation matching by phase-coded filtering have been investigated. Phase coded filtering is a technique to combine multiple patterns in one filter by assigning complex weights of unit magnitude to the individual patterns and summing them up in a composite filter. Several of the proposed composite filters are based on this idea, such as the Circular Harmonic Component (CHC) filters and the Linear Phase Coefficient Composite (LPCC) filters. In particular, ways to improve the performance of the LPCC(1) filter have been examined by assigning the complex weights to the individual patterns in a non-random manner so as to maximize the SNR of the filter with respect to the individual patterns. The trade-off between speed-up (the number of patterns combined in a filter) and unreliability (the number of resulting false matches) of the composite filter was examined by performing experiments on a database of 100 to 1,000 edge images from the aerial domain. Results indicated that for binary patterns with point densities of about 0.05, more than 20 patterns can be safely combined in the optimized LPCC(1) filter; this represents a speed-up of an order of a magnitude over the brute force approach of matching the individual patterns.

Given any two views of some unknown textured opaque quadric surface in 3D, there often exists a finite number of corresponding points across the two views that uniquely determine all other correspondences coming from points on the quadric. Identification of these points defines a "nominal" quadratic transformation that can be used in practice to facilitate the process of achieving full point-to-point correspondence between two grey-level images of the same (arbitrary) object.

6.7 Parallel image analysis [23]

Single instruction stream, single data stream (SIMD) processor array machines are popular in practical parallel computing. Such machines differ from one another considerably in the level of autonomy provided to each processing element (PE) of the array. An understanding of the levels of autonomy provided by the architectures is important in the design of efficient algorithms for them. SIMD architectures can be classified into six categories differing in key aspects, such as the selection of the instructions to be executed, operands for the instructions, and the source/destination of communications.

The data parallel model of computation used in processor arrays exploits the parallelism in the data by simultaneously processing multiple data elements (in image analysis, these elements are pixels). This is done by assigning one PE to each data element. This scheme does not make efficient use of the processor array when processing relatively small data structures. A technique of *data replication* has been developed that combines operation parallelism with data parallelism to process small data structures efficiently on large processor arrays. It decomposes the main operation into suboperations that are performed simultaneously on separate copies of the data structure. The autonomy of the individual PEs is critical to this decomposition. Replicated data algorithms have been developed for several low level image operations, such as histogramming, convolution, and rank order filtering. Additionally, a way of constructing a replicated data algorithm for an operation automatically from an

image algebra expression has been developed, thus demonstrating the generality of the approach. A replicated data algorithm also has been devised to compute single source shortest paths on general graphs, demonstrating the applicability of the approach beyond pixel-level image analysis. The speed-up performance of the algorithms has been analyzed on various interconnection networks to determine the conditions under which the technique resulted in a speedup. Implementation of the algorithms on a Connection Machine CM-2 and a MasPar MP-1 yielded impressive speedups.

A parallel search scheme also has been developed for the model-based interpretation of aerial images under a focus-of-attention paradigm; this scheme has been implemented on a CM-2. Candidate-objects are generated as connected combinations of the connected components of the image and matched against the model by checking if the parameters computed from the region satisfy the model constraints. This process is posed as a search in the space of combinations of connected components with the finding of an (optimally) successful region as the goal. The implementation exploits parallelism at multiple levels by parallelizing control tasks such as the management of the open list. The level of processor autonomy and other details of the architecture play important roles in the search scheme.

6.8 Camera motion control [34]

A hardware and software system has been developed for precise positioning and motion control of a CCD camera. The purpose of this system is to provide images, or sequences of images, with known ground truth. The system allows the testing, evaluation, and validation of various algorithms concerned with the recovery of the structure of imaged scenes and of relative three-dimensional motion from a sequence of images.

The particular focus of this work is the system software, which includes a Graphical User Interface (GUI). A key objective that drove the software design process was to have the GUI be straightforward and of minimal complexity, despite the underlying complexity of the program. And although simplicity in an interface tends to decrease a system's functionality, an attempt was made to afford the user maximum control over the camera's motion.

A self-contained user's manual, and a programmer's manual, which will allow the system to be modified in response to future needs, have been prepared; they include a general description of the system, as well as the objectives and constraints that led to its final configuration.

6.9 Function-based object recognition [29]

An approach to function-based object recognition has been developed that reasons about the functionalities of an object's intuitive parts. The popular "recognition by parts" shape recognition framework has been extended to support "recognition by functional parts," by combining a set of functional primitives and their relations with a set of abstract volumetric shape primitives and their relations. Previous approaches have relied on more global object features, often ignoring the problem of object segmentation, thereby restricting themselves to range images of unoccluded scenes. In fact, the necessary shape primitives and relations can easily be recovered from superquadric ellipsoids which, in turn, can be recovered from either

range or intensity images of occluded scenes. The framework supports both unexpected (bottom-up) object recognition and expected (top-down) object recognition. The approach has been demonstrated on a simple domain by recognizing a restricted class of hand-tools from 2D images.

7 Bibliography of reports under the Contract

1. Arya, Sunil, "Nearest Neighbor Searching and Applications." CAR-TR-777, CS-TR-3490, June 1995.

2. Black, Michael J. and Yaser Yacoob, "Tracking and Recognizing Rigid and Non-Rigid Facial Motions Using Local Parametric Models of Image Motion." CAR-TR-756, CS-TR-3401, January 1995.

3. Burlina, Philippe and Rama Chellappa, "Time-to-X: Analysis of Motion through Temporal Parameters." CAR-TR-726, CS-TR-3320, July 1994.

4. Burlina, Philippe and Rama Chellappa, "Generalized Spectral Methods for the Analysis of Spatio-Temporal Visual Information." CAR-TR-727, CS-TR-3325, July 1994.

5. Chellappa, R., C.L. Wilson, S. Sirohey and C.S. Barnes, "Human and Machine Recognition of Faces: A Survey." CAR-TR-731, CS-TR-3339, August 1994.

6. Chen, Ling Tony and Larry S. Davis, "Parallel Stereo Matching Using Dynamic Programming." CAR-TR-620, CS-TR-2884, April 1992.

7. Doermann, David S., "Document Image Understanding: Integrating Recovery and Interpretation." CAR-TR-662, CS-TR-3056, April 1993.

8. Fermüller, Cornelia and Yiannis Aloimonos, "Qualitative Egomotion." CAR-TR-629, CS-TR-2915, June 1992.

9. Fermüller, Cornelia and Yiannis Aloimonos, "The Role of Fixation in Visual Motion Analysis." CAR-TR-647, CS-TR-2989, November 1992.

10. Fermüller, Cornelia and Yiannis Aloimonos, "Vision and Action." CAR-TR-722, CS-TR-3305, June 1994.

11. Fermüller, Cornelia and Yiannis Aloimonos, "On the Geometry of Visual Correspondence." CAR-TR-732, CS-TR-3341, July 1994.

12. Fermüller, Cornelia and Yiannis Aloimonos, "Perception of 3D Motion Through Patterns of Visual Motion." CAR-TR-774, CS-TR-3484, May 1995.

13. Gavrila, D.M. and L.S. Davis, "Fast Correlation Matching in Large (Edge) Image Databases." CAR-TR-730, CS-TR-3334, August 1994.

14. Harwood, David, Timo Ojala, Matti Pietikäinen, Shalom Kelman and Larry S. Davis, "Texture Classification by Center-Symmetric Auto-Correlation, Using Kullback Discrimination of Distributions." CAR-TR-678, CS-TR-3099, July 1993.

15. Hervé, Jean-Yves, "Navigational Vision." CAR-TR-669, CS-TR-3065, May 1993.

16. Hervé, Jean-Yves, "Learning Hand/Eye Coordination by an Active Observer Part I: Organizing Centers." CAR-TR-725, CS-TR-3319, July 1994.

17. Huang, Liuqing and Yiannis Aloimonos, "The Geometry of Visual Interception." CAR-TR-622, CS-TR-2893, April 1992.

18. Kanatani, Kenichi, "Statistical Reliability of 3-D Interpretation from Images." CAR-TR-615, CS-TR-2873, April 1992.

19. Kanatani, Kenichi, "Computation of 3-D Motion from Images." CAR-TR-617, CS-TR-2879, April 1992.

20. Khuller, Samir, Balaji Raghavachari and Azriel Rosenfeld, "Localization in Graphs." CAR-TR-728, UMIACS-TR-94-92, CS-TR-3326, July 1994.

21. Khuller, Samir, Ehud Rivlin and Azriel Rosenfeld, "Graphbots: Robot Mobility in Discrete Spaces." CAR-TR-738, CS-TR-3356, September 1994.

22. Kwon, Oh-Jin, "Region Adaptive Image Coding." CAR-TR-745, CS-TR-3378, November 1994.

23. Narayanan, P.J., "Effective Use of SIMD Machines for Image Analysis." CAR-TR-635, CS-TR-2945, August 1992.

24. Netanyahu, Nathan S. and Isaac Weiss, "Analytic Line Fitting in the Presence of Uniform Random Noise." CAR-TR-783, CS-TR-3508, August 1995.

25. Oberkampf, Denis, Daniel F. DeMenthon and Larry S. Davis, "Iterative Pose Estimation Using Coplanar Feature Points." CAR-TR-677, CS-TR-3098, July 1993.

26. Rivlin, Ehud and Ronen Basri, "Localization and Positioning Using Combinations of Model Views." CAR-TR-631, CS-TR-2926, July 1992.

27. Rivlin, Ehud and Isaac Weiss, "Local Invariants for Recognition." CAR-TR-644, CS-TR-2977, October 1992.

28. Rivlin, Ehud and Isaac Weiss, "Recognizing Objects Using Deformation Invariants." CAR-TR-660, CS-TR-3041, March 1993.

29. Rivlin, Ehud, Dickinson, Sven J. and Azriel Rosenfeld, "Recognition by Functional Parts." CAR-TR-703, CS-TR-3222, February 1994.

30. Rivlin, Ehud and Azriel Rosenfeld, "Navigational Functionalities." CAR-TR-733, CS-TR-3343, August 1994.

31. Rosenblum, Mark and Larry S. Davis, "The Use of a Radial Basis Function Network for Visual Autonomous Road Following." CAR-TR-666, CS-TR-3062, May 1993.

32. Rosenblum, Mark, Yaser Yacoob and Larry S. Davis, "Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture." CAR-TR-721, CS-TR-3304, June 1994.

33. Shashua, Amnon and Sebastian Toelg, "The Quadric Reference Surface: Applications in Registering Views of Complex 3D Objects." CAR-TR-702, CS-TR-3220, February 1994.

34. Sickels, Stephen J., "A Virtual Instrument Interface for Camera Motion Control." CAR-TR-707, CS-TR-3269, May 1994.

35. Singh, Hemant and Rama Chellappa, "An Improved Shape from Shading Algorithm." CAR-TR-700, CS-TR-3218, February 1994.

36. Sirohey, Saad Ahmed, "Human Face Segmentation and Identification." CAR-TR-695, CS-TR-3176, November 1993.

37. Teng, Yang-pin Ansel, "Parallel Processing of Geometric Structures: Visibility and Triangulation Algorithms." CAR-TR-680, CS-TR-3120, August 1993.

38. Teng, Y. Ansel and Larry S. Davis, "Visibility Analysis on Digital Terrain Models and Its Parallel Implementation." CAR-TR-625, CS-TR-2900, May 1992.

39. Toelg, Sebastian, "On the Finite Kinematics of Visual Fixation." CAR-TR-736, CS-TR-3351, September 1994.

40. Toelg, Sebastian and Tomaso Poggio, "Towards an Example-Based Image Compression Architecture for Video-Conferencing." CAR-TR-723, CS-TR-3309, July 1994.

41. Waksman, Adlai and Azriel Rosenfeld, "Sparse, Opaque Three-Dimensional Texture, 2: Foliate Patterns." CAR-TR-637, CS-TR-2951, September 1992.

42. Waksman, Adlai and Azriel Rosenfeld, "Sparse, Opaque Three-Dimensional Texture, 2a: Visibility." CAR-TR-729, CS-TR-3333, July 1994.

43. Waksman, Adlai and Azriel Rosenfeld, "Sparse, Opaque Three-Dimensional Texture, 2b: Photometry." CAR-TR-740, CS-TR-3363, October 1994.

44. Wasson, Douglas D., "An Eigenspace Approach for the Enhancement of Fingerprints with Applications to Minutiae Detection." CAR-TR-709, CS-TR-3271, May 1994.

45. Weiss, Isaac, "Local Projective and Affine Invariants." CAR-TR-612, CS-TR-2870, April 1992.

46. Weiss, Isaac, "Geometric Invariants and Object Recognition." CAR-TR-632, CS-TR-2942, August 1992.

47. Weiss, Isaac, "Physics-Like Invariants for Vision." CAR-TR-752, CS-TR-3389, December 1994.

48. Weiss, Isaac, Walid G. Aref, Ehud Rivlin and Hanan Samet, "Geometric Invariants for Image Databases." CAR-TR-667, CS-TR-3063, May 1993.

49. Wu, Ting-Hu, "Estimation of Motion and Structure from Long Noisy Image Sequences." CAR-TR-689, CS-TR-3146, October 1993.

50. Wu, Ting-Hu and Rama Chellappa, "Experiments on Estimating Motion and Structure Parameters Using Long Monocular Image Sequences." CAR-TR-640, CS-TR-2969, October 1992.

51. Wu, Ting-Hu and Rama Chellappa, "Stereoscopic Recovery of Egomotion and Environmental Structure: Models, Uniqueness and Experiments." CAR-TR-646, CS-TR-2988, November 1992.

52. Yacoob, Yaser and Larry S. Davis, "Early Vision Processing Using a Multi-Stage Diffusion Process." CAR-TR-633, CS-TR-2943, August 1992.

53. Yacoob, Yaser and Larry S. Davis, "Qualitative Labeling of Human Face Components from Range Data." CAR-TR-642, CS-TR-2971, October 1992.

54. Yacoob, Yaser and Larry S. Davis, "Recognizing Human Facial Expression." CAR-TR-706, CS-TR-3265, May 1994.

55. Yacoob, Yaser, Larry S. Davis and Hanan Samet, "Scale Space Interpolation and Metamorphosis." CAR-TR-654, CS-TR-3016, January 1993.

56. Zheng, Qinfen and Rama Chellappa, "Automatic Feature Point Extraction and Tracking in Image Sequences for Arbitrary Camera Motion." CAR-TR-628, CS-TR-2911, June 1992.